CLAIMS

1. A method of obtaining a plurality of biallelic markers comprising the steps of:

obtaining a nucleic acid library comprising a plurality of genomic DNA fragments comprising the full genome or a portion thereof;

determining the order of said plurality of genomic DNA fragments in the genome;

determining the sequence of selected regions of said plurality of genomic DNA fragments; and

identifying nucleotides in said plurality of genomic DNA fragments which vary between individuals, thereby defining a set of biallelic markers.

2. The method of Claim 1, further comprising selecting a minimally overlapping set of genomic fragments from said nucleic acid library.

3. The method of Claims 1 or 2, further comprising identifying one biallelic marker per genomic DNA fragment.

4. The method of Claims 1 or 2, further comprising identifying two or more biallelic markers per genomic DNA fragment.

5. The method of Claim 1, further comprising detecting a set of biallelic markers having a desired average heterozygosity rate.

6. The method of Claims 1 or 5, further comprising selecting biallelic markers having a heterozygosity rate of at least about 0.18.

7. The method of Claims 1 or 5, further comprising selecting biallelic markers having a heterozygosity rate of at least about 0.32.

8. The method of Claims 1 or 5, further comprising selecting biallelic markers having a heterozygosity rate of at least about 0.42.

9. The method of Claim 1, wherein said identifying step comprises identifying at least about 20,000 biallelic markers.

10. The method of Claim 1, wherein said biallelic markers are separated from one another by an average distance of 10 kb - 200 kb.

11. The method of Claim 1, wherein said biallelic markers are separated from one another by an average distance of 25 kb - 50 kb.

12. The method of Claim 1, wherein the step of determining the sequence of selected regions of said plurality of genomic DNA fragments comprises inserting fragments of said plurality of genomic DNA fragments into a vector to generate a plurality of subclones and

determining the sequence of a region of the inserts in said plurality of subclones or a subset thereof.

13. The method of Claim 12, wherein said step of determining the sequence of a region of said inserts or a subset thereof comprises determining the sequence of one or both end regions of said inserts or a subset thereof.

14. The method of Claim 1, wherein a set of about 10,000 to about 30,000 genomic DNA inserts with an average size between 100 kb and 300 kb are ordered.

15. The method of Claim 1, wherein said identifying step comprises identifying between 1 and 6 biallelic markers per genomic DNA fragment.

16. The method of Claim 1, wherein said identifying step comprises identifying an average of 3 biallelic markers per genomic DNA insert.

17. The method of Claim 1, wherein said genomic DNA fragments are in a Bacterial Artificial Chromosome.

18. The method of Claim 1, further comprising determining the position of said biallelic markers along the genome or a portion thereof.

19. The method of Claim 1, further comprising obtaining pluralities of biallelic markers such that each marker is in linkage disequilibrium with at least one of identified markers.

20. The method of Claim 1, wherein said portion of the genome comprises at least 200 kb of contiguous genomic DNA.

21. The method of Claim 1, wherein said portion of the genome comprises at least 2 Mb of contiguous genomic DNA.

22. The method of Claim 1, wherein said portion of the genome comprises at least 20 Mb of contiguous genomic DNA.

23. The method of Claim 1, further comprising the step of identifying one or more groups of biallelic markers which are in proximity to one another in the genome.

24. The method of Claim 23, wherein the biallelic markers in each of these groups are located within a genomic region spanning from 1 to 5 kb.

25. The method of Claim 23, wherein the biallelic markers in each of these groups are located within a genomic region spanning from 5 kb to 1 Mb.

26. The method of Claim 23, wherein the biallelic markers in each of these groups are located within a genomic region spanning more than 1 Mb.

27. A set of biallelic markers obtained by the method of Claim 1, wherein the markers in said set are on average evenly spaced over the full genome or a portion thereof.

28. The set of biallelic markers of Claim 27, wherein the markers in said set are ordered relative to one another.

29. The set of biallelic markers according to Claim 27 or Claim 28, wherein the markers in said set have a known genomic position.

30. The set of biallelic markers of Claim 27, wherein said biallelic markers are separated from one another by an average distance of 100 to 150 kb.

31. The set of biallelic markers of Claim 27, wherein said biallelic markers are separated from one another by an average distance of 25 to 50 kb.

32. The set of biallelic markers of Claim 27, wherein said biallelic markers are separated from one another by an average distance of 10 to 200 kb.

33. The set of biallelic markers of Claim 27, wherein said biallelic markers have a heterozygosity rate of at least about 0.18.

34. The set of biallelic markers of Claim 27, wherein said biallelic markers have a heterozygosity rate of at least about 0.32.

35. The set of biallelic markers of Claim 27, wherein said biallelic markers have a heterozygosity rate of at least about 0.42.

36. A map comprising an ordered array of at least 20,000 biallelic markers obtained by the method of Claim 1.

37. A method of identifying one or more biallelic markers associated with a detectable trait comprising the steps of:

determining the frequencies of each allele of said one or more biallelic markers obtained by the method of claim 1 in individuals who express said detectable trait and individuals who do not express said detectable trait; and

identifying one or more alleles of said one or more biallelic markers which are statistically associated with the expression of said detectable trait.

38. A method of identifying a haplotype associated with a trait comprising the steps of:

obtaining nucleic acid samples from trait positive and trait negative individuals;

determining the frequencies of the alleles of each member of a group of biallelic markers obtained by the method of claim 1 located in proximity to one another in the genome in said nucleic acid samples; and

identifying a plurality of alleles of biallelic markers having a statistically significant association with said trait.

39. The method of Claim 38, wherein the biallelic markers in each of these groups are located within a genomic region spanning from 1 to 5 kb.

40. The method of Claim 38, wherein the biallelic markers in each of these groups are located within a genomic region spanning from 5 kb to 1 Mb.

41. The method of Claim 38, wherein the biallelic markers in each of these groups are located within a genomic region spanning more than 1 Mb.

42. A method of identifying one or more biallelic markers associated with a detectable trait comprising the steps of :

selecting a gene in which mutations result in a detectable trait or a gene suspected of being associated with a detectable trait; and

identifying one or more biallelic markers obtained by the method of Claim 1 within the genomic region harboring said gene which are associated with said detectable trait.

43. The method of Claim 42, wherein said identifying step comprises:

determining the frequencies of said one or more biallelic markers in individuals who express said detectable trait and individuals who do not express said detectable trait; and

identifying one or more biallelic markers which are statistically associated with the expression of said detectable trait.

44. An array of nucleic acids fixed to a support, said nucleic acids comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, of one or more biallelic markers obtained by the method of Claim 1.

45. An array of nucleic acids fixed to a support, said nucleic acids comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, of one or more groups of biallelic markers obtained by the method of Claim 1 known to be located in proximity to one another in the genome.

46. An array of nucleic acids fixed to a support, said nucleic acids comprising amplification primers for generating an amplification product comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, of one or more groups of biallelic markers obtained by the method of Claim 1 known to be located in proximity to one another in the genome.

47. An array of nucleic acids fixed to a support, said nucleic nucleic acids comprising one or more microsequencing primers for determining the identity of the polymorphic bases of one or more groups of biallelic markers obtained by the method of Claim 1 known to be located in proximity to one another in the genome.

48. An array of nucleic acids fixed to a support, wherein said nucleic acids are complementary to one or more microsequencing primers for determining the identities of the polymorphic bases of one or more biallelic markers obtained by the method of Claim 1 known to be located in proximity to one another in the genome.

AMENDED SHEET

49. The array of any one of Claims 45 to 48, wherein the members of each of said one or more groups of biallelic markers are located in physical proximity to one another on said support.

50. The array of any one of Claims 45 to 48, wherein the biallelic markers in each of these groups are located within a genomic region spanning from 1 to 5 kb.

51. The array of any one of Claims 45 to 48, wherein the biallelic markers in each of these groups are located within a genomic region spanning from 5 kb to 1 Mb.

52. The array of any one of Claims 45 to 48, wherein the biallelic markers in each of these groups are located within a genomic region spanning more than 1 Mb.

53. The array of any one of Claims 45 to 48, wherein each group of biallelic markers comprises at least 3 biallelic markers.

54. The array of any one of Claims 45 to 48, wherein each group of biallelic markers comprises at least 20 biallelic markers.

55. A method for determining whether an individual is at risk of developing a detectable trait or suffers from a detectable trait associated with said trait comprising the steps of:

obtaining a nucleic acid sample from said individual;

screening said nucleic acid sample with one or more biallelic markers obtained by the method of Claim 1; and

determining whether said nucleic acid sample contains one or more of biallelic markers statistically associated with said detectable trait.

56. The method of Claim 55, wherein said biallelic markers were obtained by the method of Claim 37.

57. The method of Claim 55, wherein said biallelic markers were obtained by the method of Claim 42.

58. A method of using a drug comprising:

obtaining a nucleic acid sample from an individual;

determining the identity of the polymorphic base of one or more biallelic markers obtained by the method of Claim 1 which is associated with a positive response to treatment with said drug or one or more biallelic markers obtained by the method of Claim 1 which is associated with a negative response to treatment with said drug; and

administering said drug to said individual if said nucleic acid sample contains one or more biallelic markers associated with a positive response to treatment with said drug or if said nucleic acid sample lacks one or more biallelic markers associated with a negative response to said drug.

59. The method of Claim 58, wherein said determining step comprises determining the identity of the polymorphic base of one or more biallelic markers obtained by the method of Claim 37 which is associated with a positive response to treatment with said drug or one or more biallelic markers obtained by the method of Claim 37 which is associated with a negative response to treatment with said drug.

60. The method of Claim 58, wherein said determining step comprises determining the identity of the polymorphic base of one or more biallelic markers obtained by the method of Claim 42 which is associated with a positive response to treatment with said drug or one or more biallelic markers obtained by the method of Claim 42 which is associated with a negative response to treatment with said drug.

61. A method of selecting an individual for inclusion in a clinical trial of a drug comprising:

obtaining a nucleic acid sample from an individual;

determining the identity of the polymorphic base of one or more biallelic markers obtained by the method of Claim 1 which is associated with a positive response to treatment with said drug or one or more biallelic markers associated with a negative response to treatment with said drug in said nucleic acid sample; and

including said individual in said clinical trial if said nucleic acid sample contains one or more biallelic markers obtained by the method of Claim 1 which is associated with a positive response to treatment with said drug or if said nucleic acid sample lacks one or more biallelic markers associated with a negative response to said drug.

62. The method of Claim 61, wherein said determining step comprises determining the identity of the polymorphic base of one or more biallelic markers obtained by the method of Claim 37 which is associated with a positive response to treatment with said drug or one or more biallelic markers obtained by the method of Claim 37 which is associated with a negative response to treatment with said drug.

63. The method of Claim 61, wherein said determining step comprises determining the identity of the polymorphic base of one or more biallelic markers obtained by the method of Claim 42 which is associated with a positive response to treatment with said drug or one or more biallelic markers obtained by the method of Claim 42 which is associated with a negative response to treatment with said drug.

64. A method of identifying a gene associated with a detectable trait comprising the steps of:

determining the frequency of each allele of one or more biallelic markers obtained by the method of Claim 1 in individuals having said detectable trait and individuals lacking said detectable trait;

identifying one or more alleles of one or more biallelic markers having a statistically significant association with said detectable trait; and

identifying a gene in linkage disequilibrium with said one or more alleles.

65. The method of Claim 64, further comprising identifying a mutation in the gene which is associated with said detectable trait.

66. A method of identifying a gene associated with a detectable trait comprising:

selecting a gene suspected of being associated with a detectable trait; and

identifying one or more biallelic markers obtained by the method of Claim 1 within the genomic region harboring said gene which are associated with said detectable trait.

67. The method of any one of Claims 37, 38, 42, 55, 64 or 66, wherein said detectable trait is selected from the group consisting of disease, drug response, drug efficacy, and drug toxicity.

68. The method of Claim 66, wherein said identifying step comprises:

determining the frequencies of said one or more biallelic markers in individuals who express said detectable trait and individuals who do not express said detectable trait; and

identifying one or more biallelic markers which are statistically associated with the expression of said detectable trait.

69. A method of identifying a haplotype associated with a trait comprising the steps of:

obtaining nucleic acid samples from trait positive and trait negative individuals;

conducting an amplification reaction on said nucleic acid samples using amplification primers capable of generating amplification products containing the polymorphic bases of a plurality of biallelic markers;

contacting one or more arrays of nucleic acids fixed to a support with said amplification products, wherein said nucleic acids fixed to a support comprise at least 8 consecutive nucleotides, including the polymorphic nucleotide, of one or more groups of biallelic markers obtained by the method of Claim 1 known to be located in proximity to one another in the genome;

determining the identities of the polymorphic bases of said amplification products; and

identifying a haplotype having a statistically significant association with said trait.

70. A method of identifying a haplotype associated with a trait comprising the steps of:

obtaining nucleic acid samples from trait positive and trait negative individuals;

conducting amplification reactions on said nucleic acid samples using amplification primers capable of generating amplification products containing the polymorphic bases of a plurality of biallelic markers;

contacting one or more arrays of nucleic acids fixed to a support with said amplification products, wherein said nucleic nucleic acids fixed to a support comprise one or more microsequencing primers for determining the identity of the polymorphic bases of one or more groups of biallelic markers obtained by the method of Claim 1 known to be located in proximity to one another in the genome;

conducting microsequencing reactions on said amplification products using microsequencing primers on said arrays, thereby generating elongated microsequencing primers comprising the polymorphic bases of said amplification products;

determining the identities of said polymorphic bases; and

identifying a haplotype having a statistically significant association with said trait.

71. A method of identifying a haplotype associated with a trait comprising the steps of:

obtaining nucleic acid samples from trait positive and trait negative individuals;

conducting amplification reactions on said nucleic acid samples uisng amplification primers which are capable of generating amplification products containing the polymorphic bases of a plurality of biallelic markers;

conducting microsequencing reactions on said nucleic acid samples, thereby generating microsequencing products containing the polymorphic bases of one or more biallelic markers at their 3' ends, said polymorphic bases being detectably labeled;

contacting one or more arrays according to Claim 48 with said microsequencing products such that said microsequencing products specifically hybridize to said nucleic acids complementary to said microsequencing primers;

determining the identities of the polymorphic bases of said microsequencing products; and

identifying a haplotype having a statistically significant association with said trait.

72. A method of identifying a haplotype associated with a trait comprising the steps of:

obtaining nucleic acid samples from trait positive and trait negative individuals;

contacting one or more arrays of nucleic acids fixed to a support with said nucleic acid sample, wherein said nucleic acids fixed to a support comprise amplification primers for generating an amplification product comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, of one or more groups of biallelic markers obtained by the method of Claim 1 known to be located in proximity to one another in the genome;

conducting an amplification reaction on said nucleic acid samples using amplification primers on said array which are capable of generating amplification products containing the polymorphic bases of a plurality of biallelic markers;

determining the identities of the polymorphic bases of said amplification products; and

identifying a haplotype having a statistically significant association with said trait.

73. A method of determining whether an individual is at risk of developing Alzheimer's disease or whether the individual suffers from Alzheimer's disease as a result of possessing the Apo E ∊4 Site A allele comprising:

obtaining a nucleic acid sample from said individual; and

determining the identity of the polymorphic base in one or more of the sequences selected from the group consisting of SEQ ID Nos. 301-305 and SEQ ID Nos. 307-311 or the sequences complementary thereto in said nucleic acid sample.

74. The method of Claim 73, further comprising determining whether said nucleic acid sample contains the sequence of SEQ ID No. 306 or the sequence complementary thereto.

75. The method of Claim 73, wherein said step of determining the identity of the polymorphic bases in one or more of the sequences selected from the group consisting of SEQ ID Nos. 301-305 and SEQ ID Nos. 307-311 or the sequences complementary thereto comprises determining whether said nucleic acid sample contains the sequence of SEQ ID No. 311 or the sequence complementary thereto.

76. The method of Claim 75, further comprising determining whether said nucleic acid sample contains the sequence of SEQ ID No. 306 or the sequence complementary thereto.

77. An isolated nucleic acid comprising a sequence selected from the group consisting of SEQ ID No. 301, SEQ ID No. 307, the sequences complementary thereto, and fragments comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, thereof.

78. An isolated nucleic acid comprising a sequence selected from the group consisting of SEQ ID No. 302, SEQ ID No. 308, the sequences complementary thereto, and fragments comprising at least 8 consecutive nucleotides thereof.

AMENDED SHEET

79. An isolated nucleic acid comprising a sequence selected from the group consisting of SEQ ID No. 303, SEQ ID No. 309, the sequences complementary thereto, and fragments comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, thereof.

80. An isolated nucleic acid comprising a sequence selected from the group consisting of SEQ ID No. 304, SEQ ID No. 310, the sequences complementary thereto, and fragments comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, thereof.

81. An isolated nucleic acid comprising a sequence selected from the group consisting of SEQ ID No. 305, SEQ ID No. 311, the sequences complementary thereto, and fragments comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, thereof.

82. An isolated nucleic acid comprising a sequence selected from the group consisting of SEQ ID Nos. 313-317, SEQ ID Nos. 319-323, and fragments comprising at least 8 consecutive nucleotides thereof.

83. An isolated nucleic acid comprising a sequence selected from the group consisting of SEQ ID Nos. 325-329, SEQ ID Nos. 331-335, the sequence complementary thereto, and fragments comprising at least 8 consecutive nucleotides thereof.

84. A set of nucleic acids comprising amplification primers for generating an amplification product comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, of one or more biallelic markers obtained by the method of Claim 1.

85. A set of nucleic acids comprising one or more microsequencing primers for determining the identity of the polymorphic base of one or more nucleic acids comprising at least 8 consecutive nucleotides, including the polymorphic nucleotide, of one or more biallelic markers obtained by the method of Claim 1.

add a1

add D2